

How can you create a residual plot by hand?

Authored by
stats writer

April 29, 2024

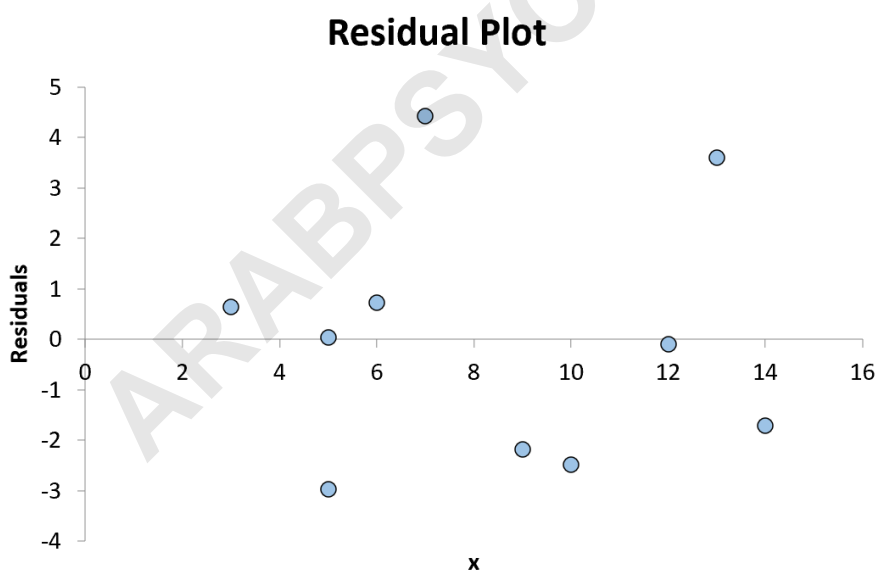
RECOMMENDED CITATION

stats writer (2024). *How can you create a residual plot by hand?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=141142>

A residual plot is a graphical representation of the differences between the observed values and the predicted values in a regression analysis. It is used to assess the validity of a linear model and to identify any patterns or outliers in the data. To create a residual plot by hand, first, calculate the predicted values for each data point using the regression equation. Then, subtract the predicted values from the observed values to obtain the residuals. Next, plot the residuals on the y-axis and the independent variable on the x-axis. The resulting plot will show the distribution of the residuals, and any patterns or outliers can be identified visually. This process can be repeated for multiple regression models, each with its own residual plot, to compare and evaluate the effectiveness of the models. Overall, creating a residual plot by hand allows for a thorough assessment of the data and can aid in the interpretation and improvement of regression models.

Create a Residual Plot by Hand

A residual plot is a type of plot that displays the values of a predictor variable in a regression model along the x-axis and the values of the along the y-axis.



This plot is used to assess whether or not the residuals in a regression model are and whether or not they

exhibit .

The following step-by-step example shows how to create a residual plot for a regression model by hand.

Step 1: Find the Predicted Values

Suppose we want to fit a regression model to the following dataset:

x	y
3	15
5	17
5	14
6	19
7	24
9	20
10	21
12	26
13	31
14	27

Using statistical software (like Excel, R, Python, SPSS, etc.) we can find that the fitted regression model is:

$$y = 10.4486 + 1.3037(x)$$

We can then use this model to predict the value of y , based on the value of x . For example, if $x = 3$, then we

would predict y to be:

$$y = 10.4486 + 1.3037(3) = 14.359$$

We can repeat this process for every in our dataset:

x	y	predicted y
3	15	14.359
5	17	16.967
5	14	16.967
6	19	18.271
7	24	19.575
9	20	22.182
10	21	23.486
12	26	26.093
13	31	27.397
14	27	28.701

Step 2: Find the Residuals

A residual for a given observation in our dataset is calculated as:

Residual = observed value - predicted value

For example, the residual of the first observation would be calculated as:

We can repeat this process for every observation in our

dataset:

x	y	predicted y	residual
3	15	14.359	0.641
5	17	16.967	0.033
5	14	16.967	-2.967
6	19	18.271	0.729
7	24	19.575	4.425
9	20	22.182	-2.182
10	21	23.486	-2.486
12	26	26.093	-0.093
13	31	27.397	3.603
14	27	28.701	-1.701

Step 3: Create the Residual Plot

Lastly, we can create a residual plot by placing the x values along the x-axis and the residual values along the y-axis.

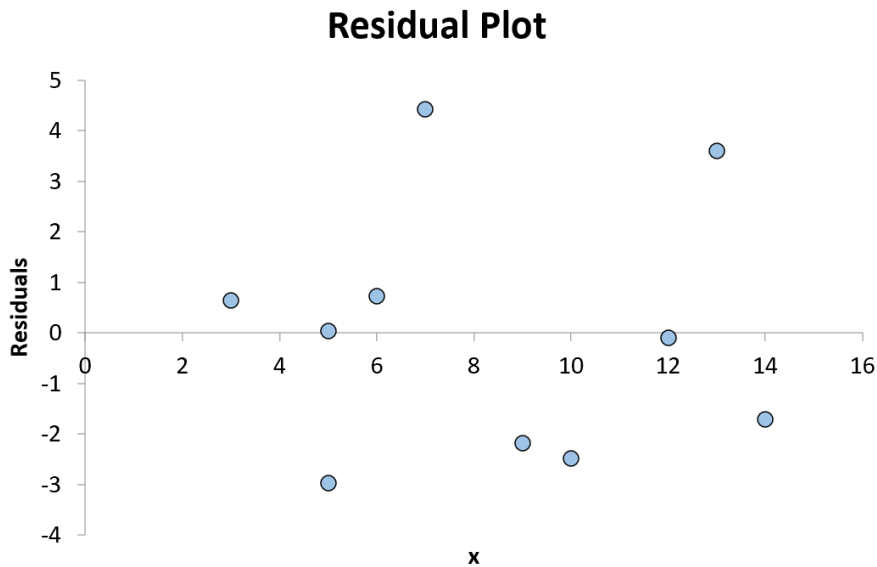
For example, the first point we'll place in our plot is (3, 0.641)

x	y	predicted y	residual
3	15	14.359	0.641
5	17	16.967	0.033
5	14	16.967	-2.967
6	19	18.271	0.729
7	24	19.575	4.425
9	20	22.182	-2.182
10	21	23.486	-2.486
12	26	26.093	-0.093
13	31	27.397	3.603
14	27	28.701	-1.701

The next point we'll place in our plot is (5, 0.033)

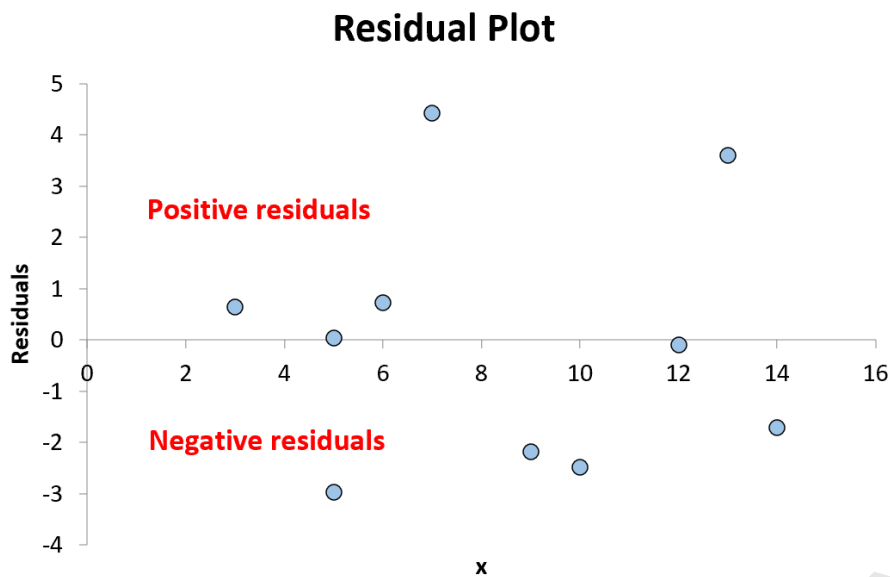
x	y	predicted y	residual
3	15	14.359	0.641
5	17	16.967	0.033
5	14	16.967	-2.967
6	19	18.271	0.729
7	24	19.575	4.425
9	20	22.182	-2.182
10	21	23.486	-2.486
12	26	26.093	-0.093
13	31	27.397	3.603
14	27	28.701	-1.701

We'll continue until we've placed all 10 pairwise combinations of x values and residual values in the plot:



Any point above zero in the plot represents a positive residual. This means the observed value for y is greater than the value predicted by the regression model.

Any point below zero represents a negative residual. This means the observed value for y is less than the value predicted by the regression model.



Since the points in the plot are randomly scattered around a residual value of 0 with no clear pattern, this indicates that the relationship between x and y is linear and a linear regression model is appropriate to use.

And since the residuals don't systematically increase or decrease as the predictor variable gets larger, this means there is not a problem with this regression model.

The following tutorials explain how to create residual plots using different statistical software: